

Automatic discrimination of core burn scars using logistic regression models

A. Bastarrika

Surveying Engineering Department, University of Basque Country, Nieves Cano 12, 01006 Vitoria-Gasteiz, Spain, aitor.bastarrika@ehu.es

E. Chuvieco

Department of Geography, University of Alcalá, Calle Colegios 2, 28801 Alcalá de Henares, Spain, emilio.chuvieco@uah.es

M.P. Martín

Institute of Economics and Geography, Spanish Council for Scientific Research (CSIC), Pinar 25-28006 Madrid, Spain, mpilar.martin@ieg.csic.es

Keywords: Burnt area mapping, logistic regression, spectral indices

ABSTRACT: Recent works have demonstrated the benefits of using a two phase methodology to improve burned area mapping from remotely sensing data. In this approach, the first phase aims at detecting the most likely burned areas (core pixels), whereas the second one improves the burned area mapping by analyzing the neighbors of previously detected pixels.

This work tries to tackle the first phase by means of several logistic regression models, using original bands and spectral indices and exploring unitemporal and multi temporal approaches. To adjust the models we used 5 pairs of Landsat TM/ETM+ data including a representative sample of burned areas in Mediterranean ecosystems. The validation of the models has been done using a Landsat scene located between Portugal and Spain with a large number of fires of different size. Visual interpretation of a color composite has been used to discriminate the fire perimeters used for validation.

Logistic regression models have shown to be an effective technique to identify burned core pixels due to their ability to identify the most appropriate variables for burned area discrimination. All proposed models achieved a correct identification of the areas affected by fires larger than 25 ha where the burned patches detection probability increased up to 95%. The models that include the Red, -NIR – and two SWIR regions have shown the most adequate, particularly the model postMIRBI-postTM54-preMIRBI-preNDVI that offered confusion rates of 0.1%, almost 100% detection probability of burned patches larger than 25 ha, and 88% for those less than 25 ha, as well as 46.7% of total burnt area detected.

1 INTRODUCTION

The spectral characterization of burnt areas is widely affected by fire intensity, residence-time and pre-fire biomass loads, all of which affect the amount of green versus scorched leaves, the proportion of ash and char, and the amount of remnant leaves. The time elapsed between fire occurrence and the acquisition of the image can also significantly alter the spectral behaviour of the burned area. Additional variations are expected for various illumination types and observation angles, especially in low resolution imagery. Consequently, spectral features of burned scars may be very diverse, and hence the automatic discrimination, especially at regional and global scales becomes very difficult.

Previous studies have tried to tackle these problems by modifying global algorithms to local and regional conditions, but yet either omission or commission errors were made, depending on whether the local thresholds match the field conditions. Moreover, these modifications limit the operational application of the proposed algorithms.

An alternative to this approach is the discrimination of burned scars in two phases: the first one aims detecting the most likely burned areas (core pixels), whereas the second one should improve the burned area mapping by analysing the neighbours of the previously detected pixels.

Based on this method (Benvenuti et al., 2000) proposed a valid approach for mapping small/medium size burned areas (larger than 1 ha) in Italy. The algorithm was tested in three test-sites in three different years (1998, 1999 and 2000) and it was based on the processing of three pairs of Landsat TM and ETM+ images, acquired before and after each fire-season. Later this methodology was applied in a semi-operative way in the ITALSCAR project (Paganini et al., 2003) to map burned areas during four consecutive years (1997-2000). Recently, Bastarrika and Chuvieco (2006) have used on this bi-phase approach in four Mediterranean sites supported by unitemporal TM and ETM+ data. This methodology was also applied to low resolution imagery. Martín (1998) proposed a similar approach to analyse

NOAA-AVHRR images, which was more recently used by (Martín et al., 2002) with MODIS data for the discrimination of large fires in the Iberian Peninsula. Fraser et al. (2002) applied this method in Canada during the 1998-2000 fire-season using VEGETATION ten day composite data. García and Chuvieco (2004) analysed SAC-C/MMRS images for burned area mapping in three large fires in Spain during the summer of 2002.

This paper focuses on the first phase of this approach. The objective of this phase is to minimize commission errors, avoiding confusion with other covers presenting similar spectral behaviour such as cloud shadows, water and topographical shade. We made use of reflectance bands and spectral indices for a set of multitemporal Landsat TM images on five sites that include a representative range of burnt areas in Mediterranean ecosystems. Several logistic regression models were developed to discriminate core burn scars.

2 METHODS

2.1 Test areas and data

The development of various logistic regression models have been accomplished using TM and ETM+ imagery of five Mediterranean areas affected by forest fires: four of them are located in Spain (*Buñol* in Valencia, *Atazar*, *Méntrida* and *Guadalajara* in the Center of the country) and another one in Greece (*Kassandra* in the Northeast of the country). These areas have been selected to take into account different land cover types that are burned in Mediterranean ecosystems. *Atazar* fire mainly burned sclerophyllous vegetation (90%), being the rest marginal agricultural areas. In *Buñol*, almost half of the burned area was also sclerophyllous (47%), and the rest transitional woodland-scrub (25%), coniferous forest (22%) and agricultural land (6%). In *Guadalajara* mainly forest mass got burned, coniferous (66%) and mixed forest (15%), together with sclerophyllous vegetation (11%) and transitional woodland-scrub (6%). Finally, fires in *Kassandra* and *Méntrida* burned mainly agricultural land (66% and 49% respectively), being transitional woodland-scrub (14% and 22%), sclerophyllous vegetation (18% and 6%) and grassland (2% and 14%) the rest. *Atazar*, *Kassandra* and *Méntrida* were medium size fires, with burned areas of 1089 ha, 1675 ha and 1900 ha, respectively, while *Guadalajara* and *Buñol* were big fires, with more than 13 000 and 23 000 ha burned, respectively.

For each of the test areas we chose a pair of Landsat TM/ETM+ images acquired before and after the fire. In all cases, the post-fire images were acquired soon after the fire, with a maximum interval of 20 days after fire extinction, whereas the pre-fire images were about the same year time.

The validation of the models was carried out in a dataset of areas located between Portugal and Spain covered by the 202-032 Landsat scene: a post ETM+ image acquired on the 5th of September of 2000 and a pre-fire image of the 30th of July of 1995; here, 417 patches smaller than 25 ha, 102 patches between 25-100 ha and 104 patches bigger than 100 ha were visually identified, with an approximately 66,910 ha total burned area. The biggest proportion is scrub and herbaceous vegetation (65%), agricultural land (20%) and forest (15%) the rest.

2.2 Geometric and atmospheric correction and reflectance

The post-fire images were geometrically corrected to UTM projection (Datum ED50 for Spain, EGSA87 for Greece and WGS84 for the validation scene between Spain and Portugal), within one pixel (RMS<1pixel) using control points and reference cartography. The pre-fire images were adjusted to the former by an image to image geometric correction to minimize misregistration errors. For the 202-032 complete scene the *itpfind* (Kennedy and Cohen, 2003) tool was used with a relative error below 0.5 pixels. The images were then converted to radiance values using sensor calibration values, and to reflectance using an atmospheric correction procedure based on the Dark Object Subtraction method (Chavez, 1996).

2.3 Development of logistic regression models

Discrimination models based on logistic regression analysis were based on both the original bands and on spectral indices that improved the discrimination of burnt areas, such as Normalized Difference Vegetation Index (NDVI), Global Environmental Monitoring Index (GEMI), Burnt Area Index (BAI), Modified Burned Area Index (BAIM), and the ratios TM7/TM4 and TM5/TM4.

Once visually delimited the burned areas (by means of TM7-TM4-TM3 colour composition), we have followed the criteria proposed by Koutsias and Karteris (1998) to extract the burned sample pixels. To get the non-burned samples we select representative samples of different land covers in each test area using as a reference the *Corine Land Cover* map. Burned samples were taken from random extractions within

the burned perimeters. 60,000 samples were obtained: 30,000 burned and 30,000 non-burned. Models have been calibrated using 50% of the samples, while the other 50% were kept for initial consistency tests.

The models were developed considering the different spectral spaces of the most common sensors used for the burned area mapping; Red-NIR spectral domain (NOAA-AVHRR/2, IRS-WiFS, ENVISAT-MERIS, SPOT-HRV), Red-NIR-SWIR domain (NOAA-AVHRR/3, SPOT-VEGETATION, IRS-AWiFS, SPOT-HRVIR) and Red-NIR-2SWIR domain -with two bands in the SWIR region- (Landsat TM y ETM+, TERRA/AQUA-MODIS, TERRA-ASTER). In this way, we have developed 10 different models using the Stepwise logistic regression method and adding the bands and indices that belong to each spectral domain considered. Some of the models have been constructed using only post-fire images and others with multi-temporal data (before and after the fire) to allow for both situations in which data before the fire is or not available.

The logistic regression models provided a probabilistic result between 0 (non-burned) and 1 (burned). Generally the threshold of 0.5 is used to classify the output variable because it represents an ideal balance between omission and commission errors. The goal of this phase was to minimize commission errors; however, it is also necessary to detect the maximum number of burned patches and this is the reason why thresholds applied are rather strict. Not all the models presented the same ability for burned patches identification; for this reason the thresholds were set up statistically at median of the burned samples used for the adjustment of models. These thresholds proved to be appropriate among all the validation data set and allowed an objective comparison between the models.

2.4 Validation

For the 203-032 scene validation perimeters were generated by visual interpretation of TM7-TM4-TM3 colour composites in the post-fire image, supported by the Portugal official perimeters. They do not provide the registration date of the fire and we have not been able to identify the fires that still do not appear in our image. This is the reason they have not been used as reference validation data.

3 RESULTS AND DISCUSSION

In general, all the models provided a correct identification of the largest fires (table 1), specially those bigger than 25 ha where the detection of the burned patches reach 95%. It is important to note that there was not cloud problems in this scene, therefore, the models based on the Red-NIR spectral regions don't have problems with the clouds shadows, one of the common problems of this spectral domain (Chuvieco et al., 2002). On this domain, the models which include BAI, specially the post-fire approach (postBAI), show large confusion with water areas. This confusion mainly arises due to the difficulty of discerning limits between water bodies and land, and also due to the construction of new water reservoirs between pre and post fire images. The postNIR-preNIR-preNDVI-postRED-postGEMI model shows similar confusion rates than de postBAI-preBAI model, but with a less significant confusion with water areas, and also between soil and scrub, although it increases the confusion with agricultural areas.

In Models that incorporate a SWIR band to the Red-NIR domain, such as the unitemporal and multitemporal BAIM models showed more confusion rate than the previous domain models, with less confusion with water areas but increased the confusion with all other land cover types, especially soil, scrub and old burned areas.

Within the models that include 2 SWIR bands, the model composed by MIRBI (postMIRBI-postTM54-preMIRBI-preNDVI) showed high confusion rates with water cover due to MIRBI index, but showed less confusion in the rest of covers than the other models, although, they detected significant less burned area than the other two models (40% vs. 70%); however, they were able to detect almost all the burn patches bigger than 25 ha, and 88% of those smaller than 25 ha. The other two models, the unitemporal postNIR-postSWIR(TM7)- postGEMI- postSWIR(TM5) and multitemporal postNIR-postSWIR(TM7)-postSWIR(TM5)- postGEMI-preNDVI-preNIR showed an effective detection of the burned areas (almost 100%), and showed less confusion with water areas (specially the unitemporal); however, they increased confusion with soil and area burned in previous years (specially the unitemporal).

4 CONCLUSIONS

The logistic regression models have demonstrated to be an effective technique to identify burn core pixels due to their sensitivity to identify the most sensible input variables for discrimination. This work incorporates different unitemporal and multitemporal models taking into account different spectral domains and thus they can be used with different type of sensors. In our validation data, the models that

include NIR and two SWIR bands presented the lowest confusion rates, specially the model postMIRBI-postTM54-preMIRBI-preNDVI, maintaining a high detection rate with almost 100% detection of burned patches larger than 25 ha, and 88% of those less than 25 ha, as well as 46.7% of burned area detected. Furthermore, it shows significant confusion with water bodies but this can be avoided using an adequate land cover data to mask those areas.

More validation work is being done to check the consistency of the results. The extension of this work would require to test its performance in different Mediterranean land covers as well as with other sensors, specially those with a low spatial and high temporal resolution that could eventually help identifying burned areas avoiding too strict thresholds.

Table I: Validation results in 202-032 scene. % detected burned patches=(patches detected)/(total of patches)x100; %detected area total = (burned correctly detected pixels/total burned pixels) x100; % confusion area total=(incorrectly burned detected pixels/total not burned pixels)x100; %confusion type cover=(incorrectly burned pixels in the cover type/ total cover type pixels) .

Model composed by	% Detected Burned patches			%Detecte d Area	%Confusion Area						
	<25 ha	25-100 ha	>100 ha		Total	Water	Soil	Agric.	Forest	Scrub	Burnt
postBAI	91 %	100 %	100 %	57,5 %	0,6 %	21,65 %	1,8 %	0,13 %	0,2 %	0,71 %	0,84 %
postBAI - preBAI	89 %	100 %	100 %	55,6 %	0,3 %	12,94 %	0,9 %	0,07 %	0,12 %	0,3 %	0,48 %
postNIR-preNIR- preNDVI- postRED- postGEMI	91 %	100 %	100 %	51,7%	0,3 %	8,83 %	0,61 %	0,15 %	0,12 %	0,19 %	0,84 %
postBAIM	76 %	95 %	100 %	39,1 %	1,2 %	7,26 %	3,68 %	0,27 %	0,47 %	2,26 %	2,21 %
postBAIM- preBAIM	76 %	95 %	100 %	38,8 %	1,1 %	7,12 %	3,47 %	0,24 %	0,44 %	2,03 %	2,03 %
postNIR- postSWIR(TM7)- postGEMI- postSWIR(TM5)	100 %	100 %	100 %	76,9 %	0,5 %	0,12 %	2,55 %	0,18 %	0,08 %	0,43 %	3,47 %
postNIR- postSWIR(TM7)- postSWIR(TM5)- postGEMI- preNDVI-preNIR	99 %	100 %	100 %	71,2 %	0,3 %	1,03 %	1,44 %	0,21 %	0,05 %	0,2 %	1,72 %
postMIRBI- postTM54- preMIRBI- preNDVI	88 %	100 %	99 %	46,7 %	0,1 %	6,97 %	0,08 %	0,01 %	0,01 %	0,02 %	0,14 %

5 ACKNOWLEDGEMENTS

Authors wish to thank Dr. Ioannis Gitas and Angela De Santis for providing the Kassandra and Guadalajara images, Felipe Verdú for his help with the management of the images and Dr. Fernando Tussel for his assistance on the development of the logistic regression models. Finally, to Eva Fernandez and Mikel Bastarrika for the assistance on the paper review.

6 REFERENCES

- Bastarrika, A. and E. Chuvieco (2006). "Cartografía del área quemada mediante crecimiento de regiones: aplicación en entornos mediterráneos con imágenes TM y ETM+." *GeoFocus* (Artículos) 6: 182-204.
- Benvenuti, M., E. Chuvieco and C. Conese (2000). A new double step methodology based on satellite image processing for forest fire mapping on the Italian territory. *EARSeL Symposium – Paris*.
- Chavez, P. S. (1996). "Image-Based Atmospheric Corrections-Revisited and Improved." *Photogrammetric Engineering&Remote Sensing* 62(9): 1025-1036.
- Chuvieco, E., M. P. Martín and A. Palacios (2002). "Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discriminations." *Int. J. Remote Sensing* 23(23): 5103-5110.
- Fraser, R. H., R. Fernandes and R. Latifovic (2002). "Multi-temporal burned area mapping using logistic regression analysis and change metrics." *IEEE Transactions on Geoscience and Remote Sensing*: 1486-1488.
- Garcia, M. and E. Chuvieco (2004). "Assessment of the potential of SAC-C/MMRS imagery for mapping burned areas in Spain." *Remote Sensing of Environment* 92(3): 414-423.
- Kennedy, R. E. and W. B. Cohen (2003). "Automated designation of tie-points for image-to-image registration." *Int. J. Remote Sensing* 24: 3467-3490.

- Koutsias, N. and M. Karteris (1998). "Logistic regression modelling of multitemporal Thematic Mapper data for burned area mapping." *Int. J. Remote Sensing* 19(18): 3499-3514.
- Martín, M. P. (1998). "Cartografía e inventario de incendios forestales en la Península Ibérica a partir de imágenes NOAA-AVHRR." Doctoral thesis, Universidad de Alcalá, Alcalá de Henares.
- Martín, M. P., R. Diaz-Delgado, E. Chuvieco and G. Ventura (2002). Burned land mapping using NOAA-AVHRR and TERRA-MODIS. *IV International Conference on Forest Fire Research*, , Luso, November, 2002. (CDRom).
- Paganini, M., O. Arino, M. Benvenuti, M. Cristaldi, M. Bordin, C. Coretti and A. Musone (2003). "ITALSCAR, a Regional Burned Forest Mapping demonstration project in Italy." *IEEE Transactions on Geoscience and Remote Sensing*: 1290-1292.